AIPL INFOSHEET
DP1 (4-97)

AIPL Data Processing Update for DRPC Workshop (April 1997)

L.M. Walton
Animal Improvement Programs Laboratory, USDA-ARS, Beltsville, MD 20705-2350
301-504-8334 (voice) 301-504-8092 (fax) rlaipl@ggpl.arsusda.gov (e-mail)

Using format 4T to interact with AIPL s editing system

Format 4T replaces formats 4 and 1, which the dairy records processing centers (DRPC's) use to submit lactation data to the Animal Improvement Programs Laboratory (AIPL) and to maintain pedigree data for grade cows. After the record cutoff for November 1997 USDA-DHIA genetic evaluations, format 4T will be used for these tasks as well as some others.

With format 4T, records can be coded such that identification (ID) changes and new or corrected lactation data can be submitted using one record by indicating a source code of D (byte 79), record-type code of R or X (byte 88), and lactation-type code of 0 to 8 (byte 126). To submit a new or corrected lactation record for pedigree and production edits, indicate a source code of D, record-type code of L, and lactation-type code of 0 to 8. Records submitted with a lactation-type code of 0 to 8 range in length from 273 to 710 bytes. Only records with a source code of D may include lactation data.

Records also can be coded in format 4T such that only pedigree maintenance is performed by indicating a source code of D; record-type code of X, P, R, D, or Y; and lactation-type code of P. Such records are 126 bytes long, and bytes 127 to 710 are ignored if present.

To correct ID errors, use a record-type code of R. This code differs from a cross-reference code (X) in that the old ID number is deleted from the AIPL data base after associating the new ID number with all existing data for the old ID number. This code should be used for correcting data entry errors. Any future data submitted with the old ID number is treated as a record submission for a new animal.

The current AIPL editing system sorts data by a cow's breed, birth year, ID number, calving date, record type (descending order), days in milk, and processing date. Although processing parents before progeny is preferred, this sort sequence is inappropriate for some situations. To give DRPC's more control over the order in which records are processed, records should be submitted to AIPL in the order the records are to be processed. If multiple files are received from a DRPC on a given day, they are processed in the order received.

As the existing AIPL editing system is moved to a workstation environment, enhancements are being made to improve the timeliness of edits. The system is being designed to handle records submitted to AIPL on a daily basis. To manage such a system efficiently, all data must be received and distributed electronically. Submitted data will be collected during the day and processed at night; error data will be placed in a directory on the AIPL system or, if feasible, returned to the DRPC electronically. To minimize data transfer time, DRPC's should submit format-4T records as variable length. An end-of-record indicator would immediately follow the last field in the record. The editing programs use a combination of the source, record-type, and lactation-type codes and the number of segments for test days to determine which data should be used in the record.

For test-day yields, AIPL would like to receive the milk yield that corresponds to the milk sample to enable accurate calculation of fat and protein yields. If electronic equipment can provide the average yield for previous days, using two of the record segments allocated for test-day information to report yield information helps facilitate calculation of component yields. One segment reports the average milk yield recorded by the electronic device with no component percentages reported; the other segment contains data for the actual test day. Although most DRPC's prefer to report only one record segment for each month, AIPL welcomes separate segments from those DRPC's willing to provide them. An alternative to submitting two segments is to adjust component percentages so that the average milk yield times the component percentage equals sample-day component yield.

Because a lactation record now can include data from multiple test days, a monthly record in progress (RIP) no longer is necessary. A record should be submitted only 1) when a cow starts or ends a lactation or changes herds, type of test, or ID; 2) when a previously rejected record needs to be corrected; or 3) at the cutoff date for genetic evaluations. Submitting data only under these circumstances will reduce the amount of data transferred as well as processing times. If more convenient, DRPC's may continue to send all RIP's on a monthly basis. However, when sending all RIP's every month, only new or corrected test days should be included in the records. Because all test-day yields are collected, both tests are needed from herds that test twice in 1 month.

If a cow changes herds during a lactation, a record for each herd should be sent. The test-day segments in each record should be limited to only those tests that were actually conducted in the herd indicated for the record.

Error reports from format-4T edits

After the record cutoff for November 1997 evaluations, error data will be distributed exclusively through the internet. Only data necessary to correct records efficiently will be returned by AIPL to the DRPC's. Feedback on an efficient method of distributing error data is appreciated. Some possible options include:

1) Return just enough ID information to identify the animal and record in question. Error segments would be appended to the end of the ID section. The current format 3 (electronic version of format 4E) is an example of this type of error reporting.

2) Return just bytes 1 to 250 of format 4T, which would exclude test-day data. Error segments would start in byte

252, and the record would range in length from 274 to 389 bytes.

3) Place error segments immediately after the last used test-day segment. The fields number of test days and number of error segments would be needed to determine where error segments began and length of the record, which would range from 297 to 412 bytes.

4) Return the format-4T record exactly as received with error segments appended to it. This record would range in length from 734 to 849 bytes.

Please note that for all four options, an end-of-record indicator would immediately follow the last filled error segment.

The artificial insemination (AI) organizations have requested more immediate access to errors and a better way of determining whether a daughter of an AI sire has problems with her lactation record. Currently, a quarterly file of error data (format 13) is provided for AI-sired daughters that have records rejected. This reporting method has been found to be difficult to work with because of untimeliness and no indication of the age of the error (no calving date, days in milk, last test date, processing date, etc.). To allow AI organizations to have better access to error information, AIPL plans to implement one or more of the following procedures:

1) Maintain an error data base that could be accessed at any time by AI organizations and DRPC's to acquire error information.

2) As lactation data are processed daily, place error data (likely one file per day) in individual directories set up on the AIPL system for AI organizations interested in facilitating the correction of records.

3) Discontinue format 13, and use the same format as used for reporting error data to the DRPC's. Production and other sensitive data could be masked. This procedure also would decrease the number of formats that AIPL needs to maintain.

Using format 1N to maintain AIPL pedigree files

Format 1N replaces the current format 1, which the breed associations use for submitting pedigree records to maintain AIPL pedigree files. Following the record cutoff for November 1997 evaluations, format 1N must be used by the breed associations when submitting pedigree maintenance records. This format includes provisions for international identification and 4-digit years for dates.

The fields in the first 99 bytes of format 1N are the same as the fields in the first 99 bytes of format 4T. Three fields included in format 1 have been excluded from format 1N: inbreeding percentage, herd code identification, and cow control number. Inbreeding percentage no longer is needed because AIPL calculates inbreeding. The other two fields had been included in format 1 for DRPC use when submitting cross-references. Two fields have been added to format 1N: processing date and registry status. At this time, processing date is not a required field, but registry status is required (although acceptable values have not yet been established).

To minimize data transfer time, breed associations should submit format-1N records as variable length. Because AIPL does not store names of females, sending data for this field is not necessary. Thus, only bytes 1 to 93 need to be reported for females. Bytes 1 to 129 should be reported for males. An end-of-record indicator would immediately follow the last field in record. Only these two record lengths are allowed. The editing programs use a combination of the source and animal gender codes to determine which data should be used in the record. Using codes to determine the length of a record requires that those fields be coded accurately to prevent misinterpretation of data.

More frequent update of herd-owner names and addresses

Currently herd-owner names and addresses are updated once a year in the spring. These names and addresses are provided in format 2N during evaluation release to help users of the elite cow report and the top 5% file contact owners of cows. Over a year's time, many of these addresses become obsolete. Therefore, AIPL will update herd-owner names and address more frequently if they are provided.

New bull evaluation format

Beginning with February 1998 evaluations, the record layout for bull evaluations provides for international ID and 4-digit years for dates. The fields also are organized into logical groupings to assist in locating data. The format number is 38 (rather than 380). Through the development process, the following fields are affected:

Fields added

- Registry status code (bytes 97-98, format 38)

Fields deleted

- Base year of evaluation (bytes 172-173, format 380)
- Date of this evaluation (bytes 49-52, format 380)
- Mean inbreeding expected from mating (bytes 209-211, format 380)
- Version code (byte 214, format 380)
- Number of lactations (goat evaluation only, bytes 291-294, format 380)
- Number of lactations for protein (goat evaluations only, bytes 378-381, format 380)

Fields changed

- All ID fields meet international ID specifications (2-byte breed code, 3-byte country code, 12-byte ID number)
- Birth date of bull accommodates a 4-digit year (bytes 89-96, format 38)
- Date bull entered AI accommodates a 4-digit year (bytes 149-154, format 38)
- Number of uniform assigned NAAB sire codes is limited

to a value of 0 to 4 because space reserved for reported NAAB codes has been limited to 4 codes (byte 165, format 38)

- All NAAB sire codes are 10 characters (3-byte AI organization code, 2-byte breed code, and 5-byte bull number) with up to 4 codes reported and the primary code reported first (bytes 166-205, format 38)
- Field format for average number of lactations per daughter for milk and fat traits is X.XX, and field length is 3 bytes. The increased precision permitted the deletion of the fields cited above (bytes 361-363, format 38)
- Field format for average number of lactation per daughter for protein traits is X.XX, and field length is 3 bytes to accommodate increased precision needed for goat evaluations (bytes 364-366)

Discontinued support for format 380S

Beginning with February 1998 evaluations, format 380S is no longer supported because of the improved speed of electronic transfer. A file based on the same selection criteria as for format 380S file is provided using format-38. The file size is estimated to be around 600,000 bytes (compared with about 220,000 bytes for the format-380S file).

Discontinued use of round reels for releasing genetic evaluation data

Beginning with May 1997 evaluations, AIPL no longer releases genetic evaluation data on round-reel magnetic tape. Use of the internet for data exchange is the preferred alternative to round-reel use. For large files that are not well suited for electronic download, 3480 and 3490 cartridges, 4-mm and 8-mm tapes, and ISO 9660 standard CD-ROM's are available.

Available on-line data

At the 1996 DRPC Workshop, a decision was made that AIPL should develop a system to provide timely access to cross-references that are entered into AIPL's data bases. A system has been implemented that created a file on the first of each month with all cross-references generated during the previous 31 days. File **usdaxrf.96200** contains all cross-references in the AIPL data bases as of July 18, 1996. Subsequent files contain monthly data. These files are available at URL **ftp://aipl.arsusda.gov/pub/usdaxrf**, as well as through AIPL's web site **http://aipl.arsusda.gov** (Downloadable Files and File Formats ▶ Monthly Files of Cross-References Added to AIPL Data Bases). The format that describes these data also is available through the AIPL web site (Downloadable Files and File Formats ▶ Description of AIPL Data Exchange Formats ▶ Other Industry Formats ▶ Cross-Reference Notification).